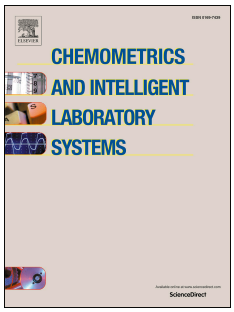
Accepted Manuscript



Efficient and robust analysis of interlaboratory studies

Jaap Molenaar, Wim P. Cofino, Paul J.J.F. Torfs

PII: S0169-7439(17)30268-X

DOI: 10.1016/j.chemolab.2018.01.003

Reference: CHEMOM 3571

To appear in: *Chemometrics and Intelligent Laboratory Systems*

Received Date: 25 April 2017

Revised Date: 11 December 2017

Accepted Date: 7 January 2018

Please cite this article as: J. Molenaar, W.P. Cofino, P.J.J.F. Torfs, Efficient and robust analysis of interlaboratory studies, *Chemometrics and Intelligent Laboratory Systems* (2018), doi: 10.1016/j.chemolab.2018.01.003.

Efficient and Robust Analysis of Interlaboratory Studies

Jaap Molenaar[1,*], Wim. P. Cofino[2], Paul J.J.F. Torfs[3]

1 Biometris, Department of Mathematical and Statistical Methods, Wageningen University, The Netherlands, jaap.molenaar@wur.nl
2 Wageningen Environmental Research (Alterra), The Netherlands, wim.cofino@wur.nl
3 Hydrology and Quantitative Water Management Group, Wageningen University, The Netherlands, paul.torfs@wur.nl

*Corresponding author

Abstract

In this paper we present the ab-initio derivation of an estimator for the mean and variance of a sample of data, such as obtained from proficiency tests. This estimator has already been used for some time in this kind of analyses, but a thorough derivation together with a detailed analysis of its properties is missing until now. The estimator uses the information contained in data including uncertainty, represented via probability density functions (pdfs). An implementation of the approach is given that can be used if the uncertainty information is not available; the so-called normal distribution approximation. The present estimation procedure is based on calculating the centroid of the ensemble of pdfs. This centroid is obtained by solving the eigenvalue problem for the so-called similarity matrix. Elements of this matrix measure the similarity (or overlap) between different pdfs in terms of the Bhattacharyya coefficient. Since evaluation of an eigenvalue problem is standard nowadays, the method is extremely fast. The first and second moments of the centroid pdf are used to obtain the mean and variance of the dataset.

The properties of the estimator are extensively analyzed. We derive its variance and show the connection between the present estimator and Principal Component Analysis. Furthermore, we study its behavior in severall limiting cases, as met in data that are very coherent or very incoherent, and check its consistency. In particular, we investigate how sensitive the estimator is for outliers, investigating its breakdown point. In the normal distribution approximation the breakdown point of the estimator is shown to be optimal, i.e., 50%.

The largest eigenvalue(s) of the similarity matrix appear(s) to provide important information. If the largest eigenvalue is close to the dimension of the matrix, this indicates that the data are very coherent, so they lie close to each other with similar uncertainties. If there are two (or more) largest eigenvalues with (nearly) equal values, this indicates that the data fall apart in two (or more) clusters.

Keywords: interlaboratory studies, proficiency tests, population characteristics, consistency, breakdown point, robust statistics

Introduction

Interlaboratory studies play an important role in analytical chemistry to assure the quality of measurements. In these studies a group of laboratories receive identical portions from a homogeneous and stable test sample for analysis. Three types of studies